1996

From Snakes to Stars, the Statistics of Collapsed Objects - II. Testing a Generic Scaling *Ansatz* for Hierarchical Clustering

Dipak Munshi[1,4], Peter Coles[1,3] and Adrian L. Melott[2],
[1] *Queen Mary and Westfield College, London E1 4NS, United Kingdom*
[2] *Department of Physics and Astronomy, University of Kansas, Lawrence, Kansas 66045, U.S.A.,*
[3] *School of Physics and Astronomy, University of Nottingham, University Park, Nottingham NG7 2RD, United Kingdom*
[4] *International School for Advanced Studies (SISSA), Via Beirut 2-4, I-34013 Trieste, Italy*

16 February 1999

ABSTRACT
We develop a diagrammatic technique to represent the multi-point cumulative probability density function (CPDF) of mass fluctuations in terms of the statistical properties of individual collapsed objects and relate this to other statistical descriptors such as cumulants, cumulant correlators and factorial moments. We use this approach to establish key scaling relations describing various measurable statistical quantities if clustering follows a simple general scaling *ansatz*, as expected in hierarchical models. We test these detailed predictions against high-resolution numerical simulations. We show that, when appropriate variables are used, the count probability distribution function (CPDF) shows clear scaling properties in the non-linear regime. We also show that analytic predictions made using the scaling model for the behaviour of the void probability function (VPF) also match the simulations very well. We generalise the results for the CPDF to the two-point (bivariate) count probability distribution function (2CPDF), and show that its behaviour in the simulations is also well described by the theoretical model, as is the bivariate void probability function (2VPF). We explore the behaviour of the bias associated with collapsed objects in limit of large separations, finding that it depends only on the intrinsic scaling parameter associated with collapsed objects and that the bias for two different objects can be expressed as a product of the individual biases of the objects. Having thus established the validity of the scaling ansatz in various different contexts, we use its consequences to develop a novel technique for correcting finite-volume effects in the estimation of multi-point statistical quantities from observational data.

Key words: Cosmology: theory – large-scale structure of the Universe – Methods: statistical

1 INTRODUCTION

The process of gravitational instability leads to the growth and collapse of initially small inhomogeneities in a self-gravitating medium. This is the basic idea behind theories for the formation of galaxies and large-scale structures in the Universe, as the small initial fluctuations seed the formation of collapsed objects that progressively cluster and merge into objects of larger size, such as galaxies and clusters. Much progress has been made during the last two decades using numerical simulations to follow the evolution of cosmological density fluctuations up to and beyond their collapse into bound structures, but it is nevertheless important to develop an analytical framework for describing this phenomenon and understanding what the simulation results.

Much of the previous analytic work in this direction has been based on an extended version of the Press-Schechter (1974) model. Such methods have been used to study the distribution of formation epochs (Bond et al. 1991; Lacey & Cole 1993; see White 1993 for a complete review), merger rates and survival times of haloes and also how the distribution of collapsed objects relates to the underlying mass distribution (Mo & White 1996; Mo et al. 1997). Tests of these predictions have shown good agreement with theoretical model predictions. Other recent methods which try to explain statistics of collapsed dark objects depend on extension of Zel'dovich approximation (Lee & Shandarin 1997, 1998; Catelan et al. 1998; Porciani et al. 1998).

A parallel approach has been to assume that statistical properties of the distribution of collapsed objects - particularly the many-body correlation functions - follow a hierarchical scaling pattern (Balian & Schaeffer 1989a,b). This kind of scaling can be used to predict, for example, the mass function of collapsed objects (Valageas & Schaeffer 1997) and their correlation properties (Bernardeau & Schaeffer 1992; Munshi et al. 1998b). These results are especially interesting because they relate directly to the statistics of the underlying mass distribution and hence to the dynamical origin of the distribution governed by the BBGKY equations.

Predictions of the behaviour of one-point statistical quantities such as the count probability distribution function (CPDF) or void probability distribution function (VPF) for the underlying mass distribution in the highly non-linear regime of strong clustering were derived using this kind of scaling *ansatz* by Balian & Schaeffer (1988, 1989). These results were later extended to the quasi-linear regime where a perturbative series expansion can be used (Beranardeau 1992, 1994a, 1994b, 1995). It emerged from these and other studies that a scaling hierarchy naturally developed by the action of gravitational clustering on Gaussian (random-phase) initial perturbations. Predictions for these one-point quantities have been tested with reasonable success against observations using various galaxy catalogs (Maurogordato & Lachieze-Rey 1987; Blanchard et al. 1990; Mourogordato et al. 1992; Bouchet et al. 1993) and numerical simulations (Bouchet & Hernquist 1992; Lucchin et. al 1994; Colombi et al. 1992, 1994, 1995, 1996; Munshi et al. 1998c), although Lucchin et al. (1994) and Colombi et al. (1996) found some deviation from scaling.

It is natural now to ask whether these properties of one-point statistical quantities can be generalised to two-point statistics, or even beyond. The two-point count probability distribution function (2CPDF) and two-point void probability function (2VPF), both defined in detail below, are particularly interesting for a number of different reasons. The first is that the 2VPF can be shown to be a generating function for the cumulant correlators (CCs) in a way which is very similar to the VPF, which acts as a generating function for cumulants. Recently it was shown that CCs can be reliably measured using galaxy catalogs (Szapudi et al. 1992; Meiksin et al. 1992; Szapudi et al 1995; Szapudi & Szalay 1997). They were also measured in numerical simulations (Munshi & Melott 1998) to test extended perturbation theory (EPT) as proposed by Colombi et al. (1996). Since it becomes increasingly difficult to measure these quantities with increasing order, it is sensible instead to study their generating function (2VPF) which carries information to all orders. The second reason is that these two-point objects are useful in the problem of estimating errors associated with the determination of one-point quantities such as the CPDF, VPF and S_N parameters (Szapudi & Colombi 1995). Although analytical results related to two-point quantities are frequently used to estimate and correct such errors, no systematic studies have yet been done to test these predictions directly against numerical simulations, partly because of the difficulty of performing numerical simulations with the very large dynamic range necessary to study these quantities. Finally, the 2CPDF carries information as to how overdense regions are biased tracers of the underlying mass distribution, so a computation of the bias implied by the scaling ansatz can provide an understanding of the origin of bias that can be compared with results from other approaches.

Before proceeding, however, we stress that one should not get too carried away by the success of this kind of scaling model. The hierarchical ansatz is, at best, a simplifying hypothesis which, admittedly seems to work very well for "realistic models" of structure formation. But no firm connection has yet been established with the microscopic physics of collisionless clustering described by BBGKY equations (Peebles 1980). Most efforts in solving the BBGKY equations have focused on particular closure schemes (Davis & Peebles 1977), the general separability of position correlations and momentum correlation in phase-space (Fry 1984; Hamilton 1988) and its stability properties (Ruamsuwan & Fry 1992; Yano & Gouda 1998). Although these precious efforts in solving the highly complicated nonlinear integro-differential BBGKY equations provide us with valuable insights, they also help underline the complications in arriving in any general solution to the problem. Our inability to solve these equations in the highly non-linear regime leaves us with undetermined parameters which can only be computed from numerical simulations, thus complicating the task of testing the scaling ansatz itself.

This paper is structured in the following way. Section 2 is devoted to a theoretical discussion of multi-point cumulants, multi-point factorial moments, multi-point factorial correlators and their various generating functions. We also develop a general approach to the treatment of finite-volume effects using these quantities. Section 3 is devoted specifically to one-point and two-point statistics. We summarize the main analytical results of Balian & Schaeffer (1989) and Bernardeau & Schaeffer (1992) and test them numerically. Section 4 contains a detailed description of our numerical simulations and to details of data analysis techniques; we also present the results of our analysis in this section. We discuss the main results and relate it to other studies in Section 5.

2 VOIDS & PROBABILITIES FROM CUMULANT CORRELATORS

The void probability function (VPF) is the probability that a randomly-placed sphere of some volume V contains no galaxies. This can be written

$$P_V(0) = \exp\left(-\frac{\phi(N_c)}{\bar{\xi}_2}\right), \tag{1}$$

in which the mean number of galaxies per cell is $\bar{N}$ and the volume-average of the two-point correlation function over the cell is

$$\bar{\xi}_2 = \frac{1}{V^2}\int\int \xi_2(\mathbf{r}_1,\mathbf{r}_2)dV_1dV_2. \tag{2}$$

The quantity $N_c = \bar{N}\bar{\xi}_2$. The VPF is a generating function of $P(N)$, the count probability distribution function (CPDF) for cells of the same size (Balian & Schaeffer 1989). Let us assume Using the generating function relationship between the CPDF and VPF we can write

$$P(N) = \frac{(-1)^N}{N!} \frac{d}{d\mu^N} \exp\left(-\frac{\phi(\mu N_c)}{\bar{\xi}_2}\right)\Bigg|_{\mu=1}, \tag{3}$$

which can also be written in the following integral form (Munshi et al. 1998c)

$$P(N) = \frac{1}{2\pi i} \int \frac{d\lambda}{\lambda^{N+1}} \exp\left(-\frac{\phi((1-\lambda)N_c)}{\bar{\xi}_2}\right). \tag{4}$$

The integral is to be evaluated along a contour around $\lambda = 0$. We also define the function $\Psi(t) = -\phi(-t)$ which can be directly related to the function $\Pi(\nu)$, the continuous analogue of $P(N)$,

$$\exp\left(\frac{\Psi(t)}{\bar{\xi}_2}\right) = \int_0^\infty d\nu \exp\left(\frac{\nu t}{N_c}\right) \Pi(\nu). \tag{5}$$

Using the definition of $P(N)$ now one can easily show that

$$P(N) = \int_0^\infty d\nu \frac{\exp(-\nu)\nu^N}{N!} \Pi(\nu). \tag{6}$$

This is the form of distribution obtained when $\Pi(\nu)$ is Poisson sampled, because it is the convolution of the function $\Pi(\nu)$ with a Poisson distribution describing the shot noise effects. This demonstrates that $\Pi(\nu)$ can be viewed as the continuous limit of $P(N)$ in the limit of large number densities. This can be seen by change of variable $\lambda = 1 + t/N_c$ in equation (3) and then taking the limit $N_c \to \infty$ and $N \to \infty$ with the ratio N/N_c remaining finite; this also gives $P(N) = \Pi(N)$.

The factorial moments of $P(N)$ can now be related to the moments of $\Pi(\nu)$ with the aid of the definition of the p-th factorial moment:

$$\langle (N_1)_p \rangle = \sum_{N=0}^{\infty} N(N-1)\dots(N-p+1)P(N) = \int_0^\infty \nu^p \Pi(\nu) d\nu. \tag{7}$$

One can similarly define the normalized factorial moments of $P(N)$ by

$$\Sigma_p = \frac{\bar{\xi}_2}{p! N_c^p} \sum N(N-1)...(N-p+1)P(N) = \frac{\bar{\xi}_2}{p! N_c^p} (N_1)_p. \tag{8}$$

One can define a generating function $\Sigma(t)$ for the Σ_p parameters using

$$\Sigma(t) = \sum_1^\infty \Sigma_p t^p \tag{9}$$

and we have the following expression connecting these two generating functions $\Psi(t)$ and $\Sigma(t)$ (Munshi et al. 1998c),

$$1 + \frac{\Sigma(t)}{\bar{\xi}_2} = \exp(\frac{\Psi(t)}{\bar{\xi}_2}), \tag{10}$$

which can also be written as

$$\Psi(t) = \bar{\xi}_2 \ln(1 + \frac{\Sigma(t)}{\bar{\xi}_2}). \tag{11}$$

It is possible to extend the method of generating functions to multi-point factorial moments. If we denote the two-point CPDF $P(N_1, N_2)$, the joint probability of finding N_1 particles in the first cell and N_2 particles in a second cell, and assume locally Poisson sampling distribution then we can relate $P(N_1, N_2)$ with its smooth counterpart $\pi(\nu_1, \nu_2)$ by:

$$P(N_1, N_2) = \int_0^\infty d\nu_1 \int_0^\infty d\nu_2 \frac{\exp(-\nu_1)\nu_1^{N_1}}{N_1!} \frac{\exp(-\nu_2)\nu_2^{N_2}}{N_2!} \Pi(\nu_1, \nu_2) \tag{12}$$

so that the two-point factorial moment becomes

$$\langle (N_1 N_2)_{pq} \rangle = \sum_{N=0}^{\infty} N_1(N_1-1)\dots(N_1-p+1)N_2(N_2-1)\dots(N_2-q+1)P(N_1,N_2) = \int_0^\infty \nu_1^p d\nu_1 \int_0^\infty \nu_2^q d\nu_2 \; \Pi(\nu_1,\nu_2) \tag{13}$$

We define the normalised two-point factorial moment by following equation.

$$\Sigma^{(2)}_{pq} = \frac{\bar{\xi}_2^2}{p!q! N_c^{p+q}} \sum N_1(N_1-1)\dots(N_1-p+1)N_2(N_2-1)\dots(N_2-q+1)P(N_1,N_2) = \frac{\bar{\xi}_2^2}{p!q! N_c^{p+q}} (N_1 N_2)_{pq} \tag{14}$$

It is to be noted that, in general, owing to statistical isotropy and homogeneity Σ_{pq} should depend only on the separation of two cells. We define the generating function for Σ_{pq} by $\Sigma^{(2)}$, i.e.

$$\Sigma^{(2)} = \sum_{p,q=1}^{\infty} \Sigma_{pq} t_1^p t_2^q \tag{15}$$

and similarly the generating function for the 2CCs by

$$\frac{\Psi^{(2)}(t_1,t_2)}{\bar{\xi}_2^2} = \sum \frac{t_1^p t_2^q}{p!q!} \frac{\langle \delta^p(x_1)\delta^q(x_2)\rangle}{\langle \delta^2(x)\rangle^{(p+q)}}. \tag{16}$$

Using similar technique as used for the case of one-point moments we can relate $\Sigma^{(2)}(t_1,t_2)$ with $\Psi^{(2)}(t_1,t_2)$:

$$1 + \frac{\Sigma(t_1)}{\bar{\xi}_2} + \frac{\Sigma(t_2)}{\bar{\xi}_2} + \frac{\Sigma^{(2)}(t_1,t_2)}{\bar{\xi}_2^2} = \exp\left(\frac{\Psi(t_1)}{\bar{\xi}_2} + \frac{\Psi(t_2)}{\bar{\xi}_2} + \frac{\Psi^{(2)}(t_1,t_2)}{\bar{\xi}_2^2}\right) \tag{17}$$

A Taylor expansion of the above equation provides us with a set of relations which can be used to measure $\Psi_{pq}^{(2)}$ once $\Sigma_{pq}^{(2)}$ is computed:

$$\Psi_{11}^{(2)} = (-1 + \Sigma_{11}^{(2)}) \tag{18}$$

$$\Psi_{12}^{(2)} = \frac{1}{\bar{\xi}_2}(1 - \Sigma_{11}^{(2)} + \bar{\xi}_2(\Sigma_{12}^{(2)} - \Sigma_2) \tag{19}$$

$$\Psi_{22}^{(2)} = \frac{1}{\bar{\xi}_2^2}(-3 + 4\Sigma_{11}^{(2)} - {\Sigma_{11}^{(2)}}^2 - 2\xi_2(\Sigma_{12}^{(2)} - 2\Sigma_2 + \Sigma_{21}^{(2)}) - 2\xi_2^2(\Sigma_2^2 - \Sigma_{22}^{(2)})) \tag{20}$$

$$\Psi_{13}^{(2)} = \frac{1}{\bar{\xi}_2^2}(-1 + \Sigma_{11}^{(2)} - \bar{\xi}_2(\Sigma_{12}^{(2)} - 2\Sigma_2) - \bar{\xi}_2\Sigma_{11}^{(2)}\Sigma_2 + \bar{\xi}_2^2(\Sigma_{13}^{(2)} - \Sigma_3)) \tag{21}$$

$$\begin{aligned}\Psi_{23}^{(2)} = \frac{1}{\bar{\xi}_2^3}&(2 + {\Sigma_{11}^{(2)}}^2 - \Sigma_{11}(3 + \bar{\xi}_2(\Sigma_{12}^{(2)} - 2\Sigma_2)) \\ &+\bar{\xi}_2(2\Sigma_{12}^{(2)} - 4\Sigma_2 + \Sigma_{21}^{(2)})) - \bar{\xi}_2^2(\Sigma_{13}^{(2)} - 2\Sigma_2^2 + \Sigma_2\Sigma_{21}^{(2)} + \Sigma_{22} - \Sigma_3) + \bar{\xi}_2^3(\Sigma_{23} - \Sigma_2\Sigma_3))\end{aligned} \tag{22}$$

In our above analysis we have used two different cells of the same size, but it is now obvious that the corresponding result for two cells of different size will read

$$1 + \frac{\Sigma(t_1)}{\bar{\xi}_2^x} + \frac{\Sigma(t_2)}{\bar{\xi}_2^y} + \frac{\Sigma^{(2)}(t_1,t_2)}{\bar{\xi}_2^x\bar{\xi}_2^y} = \exp\left(\frac{\Psi(t_1)}{\bar{\xi}_2^x} + \frac{\Psi(t_2)}{\bar{\xi}_2^y} + \frac{\Psi^{(2)}(t_1,t_2)}{\bar{\xi}_2^x\bar{\xi}_2^y}\right). \tag{23}$$

These results can be extended reasonably straightforwardly to three-point quantities:

$$1 + \frac{\Sigma(t_1)}{\bar{\xi}_2} + \frac{\Sigma(t_2)}{\bar{\xi}_2} + \frac{\Sigma(t_3)}{\bar{\xi}_2} + \frac{\Sigma^{(2)}(t_1,t_2)}{\bar{\xi}_2^2} + \frac{\Sigma^{(2)}(t_1,t_3)}{\bar{\xi}_2^2} + \frac{\Sigma^{(2)}(t_2,t_3)}{\bar{\xi}_2^2} + \frac{\Sigma^{(3)}(t_1,t_2,t_3)}{\bar{\xi}_2^3} = \exp\chi_3, \tag{24}$$

where

$$\chi_3 = \frac{\Psi(t_1)}{\bar{\xi}_2} + \frac{\Psi(t_2)}{\bar{\xi}_2} + \frac{\Psi(t_2)}{\bar{\xi}_2} + \frac{\Psi^{(2)}(t_1,t_2)}{\bar{\xi}_2^2} + \frac{\Psi^{(2)}(t_2,t_3)}{\bar{\xi}_2^2} + \frac{\Psi^{(2)}(t_1,t_3)}{\bar{\xi}_2^2} + \frac{\Psi^{(3)}(t_1,t_2,t_3)}{\bar{\xi}_2^3} \tag{25}$$

and

$$\frac{\Psi^{(3)}(t_1,t_2,t_3)}{\bar{\xi}_2^3} = \sum \frac{t_1^p t_2^q t_3^r}{p!q!r!} \frac{\langle \delta^p(x_1)\delta^q(x_2)\delta^r(x_3)\rangle}{\bar{\xi}_2^{(p+q+r)}} \tag{26}$$

The quantity $\Sigma^{(3)}$ is similarly the three-point generalisation of the two-point quantity $\Sigma^{(2)}$. Expanding this relation one can express three-point moments in terms of factorial moments of the three-point CPDF.

$$\Psi_{111}^{(3)} = (2 - 3\Sigma_{11}^{(2)} + \Sigma_{111}^{(3)}) \tag{27}$$

$$\Psi_{112}^{(3)} = \frac{1}{\bar{\xi}_2}(-3 - {\Sigma_{11}^{(2)}}^2 - \Sigma_{111}^{(3)} + \bar{\xi}_2\Sigma_{112}^{(3)} - 2\bar{\xi}_2\Sigma_{12}^{(2)} + 2\bar{\xi}_2\Sigma_2 + \Sigma_{11}^{(2)}(5 - \bar{\xi}_2\Sigma_2)) \tag{28}$$

$$\begin{aligned}\Psi_{113}^{(3)} = \frac{1}{\bar{\xi}_2^2}&(4 + 2\Sigma_{11}^2 + \Sigma_{111}^{(3)} + 4\xi_2\Sigma_{12}^{(2)} - 2\bar{\xi}_2^2\Sigma_{13}^{(2)} - 6\bar{\xi}_2\Sigma_2 - \bar{\xi}_2\Sigma_{111}^{(3)}\Sigma_2 \\ &-\bar{\xi}_2\Sigma_{112}^{(3)} + 2\bar{\xi}_2^2\Sigma_3 - \Sigma_{11}^{(2)}(7 + 2\xi_2(\Sigma_{12}^{(2)} - 3\Sigma_2) + \bar{\xi}_2^2\Sigma_3) + \bar{\xi}_2^2\Sigma_{113})\end{aligned}$$

It is possible to generalise these equations for an arbitrary number of points:

$$1 + \sum \frac{\Sigma(t_i)}{\bar{\xi}_2^2} + \sum_{\text{pairs}} \frac{\Sigma^{(2)}(t_i,t_j)}{\bar{\xi}_2^2} + \sum_{\text{triplets}} \frac{\Sigma^{(3)}(t_i,t_j,t_k)}{\bar{\xi}_2^2} + \ldots + \frac{\Sigma^{(p)}(t_i \ldots t_n)}{\bar{\xi}_2^p} = \exp\chi_p, \tag{29}$$

where

$$\chi_p = \sum \frac{\Psi(t_i)}{\bar{\xi}_2} + \sum_{\text{pairs}} \frac{\Psi^{(2)}(t_i,t_j)}{\bar{\xi}_2^2} + \sum_{\text{triplets}} \frac{\Psi^{(3)}(t_i,t_j,t_k)}{\bar{\xi}_2^3} \ldots + \frac{\Psi^{(p)}(t_1,t_2,\ldots t_k)}{\bar{\xi}_2^p}. \tag{30}$$

It is important to note that two-point quantities depend only on the separation of two cells, but the multiple-point correlators and their generating functions depend on the geometrical configuration of different cells. We have used the following definition for multiple-point cumulant correlators:

$$\frac{\Psi^{(k)}(t_1,t_2,\ldots t_k)}{\bar{\xi}_2^k} = \sum\left(\frac{t_1^p t_2^q t_3^r \ldots t_k^s}{p!q!r!\ldots s!}\right)\frac{\langle\delta^p(x_1)\delta^q(x_2)\delta^r(x_3)\ldots\delta^s(x_k)\rangle}{\bar{\xi}_2^{(p+q+r+\ldots+s)}} \tag{31}$$

Inverting this relation, we can express $\Psi^{(k)}(t_1,t_2,\ldots,t_k)$ in terms of the generating function for factorial moments, which can be measured directly from simulations:

$$\begin{aligned}\Psi^{(k)}(t_1,t_2,\ldots,t_k) &= \ln M_k(t_1,t_2,\ldots,t_k) - \sum_{(k-1)\text{tuplets}} \ln M_{k-1}(t_1,t_2,\ldots,t_{k-1}) + \sum_{(k-2)\text{tuplets}} M_{k-2}(t_1,t_2,\ldots,t_{k-2}) + \\ &\ldots + (-1)^{(k-2)}\sum_{\text{pairs}}\ln M_2(t_i,t_j) + (-1)^{(k-1)}\sum_{\text{singlets}}\ln M(t_i)\end{aligned} \tag{32}$$

where we have defined

$$M(t_1) = 1 + \frac{\Sigma(t_1)}{\xi_2} \tag{33}$$

$$M_2(t_1,t_2) = 1 + \frac{\Sigma(t_1)}{\xi_2} + \frac{\Sigma(t_2)}{\xi_2} + \frac{\Sigma^{(2)}(t_1,t_2)}{\xi_2^2}. \tag{34}$$

It is also possible to work with factorial correlators instead of the factorial moments defined above by extending the two-point factorial correlators introduced by Szapudi & Szalay (1995) to multiple points. If we denote the k^{th} order factorial moments in different cells by $(N_i)_k$ then we can write the generating function of the two-point factorial correlators W_{kl} by $W(t_1,t_2)$, the generating function of the three-point factorial correlator W_{klm} by $W(t_1,t_2,t_3)$, and so on. In general $W(t_1,\ldots,t_s)$ denotes the generating function for s-point factorial correlators $W_{k\ldots r}$:

$$W^{(1)}(t_1) = 1 + \sum_{k=1}^{\infty}\frac{1}{k!}\left(\frac{t_1}{\bar{N}}\right)^k\langle(N_1)_k\rangle \tag{35}$$

$$W^{(2)}(t_1,t_2) = \sum_{k=1,l=1}^{\infty}\frac{1}{k!l!}\left(\frac{t_1}{\bar{N}}\right)^k\left(\frac{t_2}{\bar{N}}\right)^l\langle((N_1)_k - \langle(N_1)_k\rangle)((N_2)_l - \langle(N_2)_l\rangle)\rangle \tag{36}$$

$$W^{(3)}(t_1,t_2,t_3) = \sum_{k=1,l=1,m=1}^{\infty}\frac{1}{k!l!m!}\left(\frac{t_1}{\bar{N}}\right)^k\left(\frac{t_2}{\bar{N}}\right)^l\left(\frac{t_3}{\bar{N}}\right)^m\langle((N_1)_k - \langle(N_1)_k\rangle)((N_2)_l - \langle(N_2)_l\rangle)((N_2)_m - \langle(N_2)_m\rangle)\rangle \tag{37}$$

$$W^{(s)}(t_1,\ldots,t_s) = \sum_{k=1,\ldots,r=1}^{\infty}\frac{1}{k!\ldots r!}\left(\frac{t_1}{\bar{N}}\right)^k\ldots\left(\frac{t_s}{\bar{N}}\right)^r\langle((N_1)_k - \langle(N_1)_k\rangle)\ldots((N_s)_r - \langle(N_s)_r\rangle)\rangle \tag{38}$$

Now it is possible to link the basis function $W^{(n)}$ for computing multi-point cumulant correlators with the other set of basis functions $\Sigma^{(n)}$, which carries equivalent information.

$$W\left(\frac{t_1}{\bar{\xi}_2}\right) = 1 + \frac{\Sigma(t)}{\bar{\xi}_2} \tag{39}$$

$$W^{(2)}\left(\frac{t_1}{\bar{\xi}_2},\frac{t_2}{\bar{\xi}_2}\right) = \frac{1}{\xi_2^2}\Sigma^{(2)}(t_1,t_2) - \frac{1}{\xi_2}\Sigma(t_1)\frac{1}{\xi_2}\Sigma(t_2) \tag{40}$$

$$\begin{aligned}W^{(3)}\left(\frac{t_1}{\bar{\xi}_2},\frac{t_2}{\bar{\xi}_2},\frac{t_3}{\bar{\xi}_2}\right) &= \frac{1}{\xi_2^3}\Sigma^{(3)}(t_1,t_2,t_3) + 2\frac{1}{\xi_2}\Sigma(t_1)\frac{1}{\xi_2}\Sigma(t_2)\frac{1}{\xi_2}\Sigma(t_3) \\ &\quad -\frac{1}{\xi_2}\Sigma(t_1)\frac{1}{\xi_2^2}\Sigma^{(2)}(t_2,t_3) - \frac{1}{\xi_2}\Sigma(t_2)\frac{1}{\xi_2^2}\Sigma^{(2)}(t_1,t_3) - \frac{1}{\xi_2}\Sigma(t_3)\frac{1}{\xi_2^2}\Sigma^{(2)}(t_1,t_2)\end{aligned} \tag{41}$$

Relating multi-point factorial moments with generating functions of multiple-point cumulant correlators (defined earlier) we can write:

$$\exp(\chi_1) = \exp\left(\frac{\Psi(t_1)}{\xi_2}\right) = W\left(\frac{t}{\bar{\xi}_2}\right); \tag{42}$$

$$\exp(\chi_2) = \exp\left(\frac{\Psi(t_1)}{\xi_2} + \frac{\Psi(t_2)}{\xi_2} + \frac{\Psi^{(2)}(t_1,t_2)}{\xi_2^2}\right) = W^{(2)}\left(\frac{t_1}{\bar{\xi}_2},\frac{t_2}{\bar{\xi}_2}\right) + W\left(\frac{t_1}{\bar{\xi}_2}\right)W\left(\frac{t_2}{\bar{\xi}_2}\right) \tag{43}$$

$$\exp(\chi_3) = W^{(3)}\left(\frac{t_1}{\bar{\xi}_2},\frac{t_2}{\bar{\xi}_2},\frac{t_3}{\bar{\xi}_2}\right) \tag{44}$$

$$+W^{(2)}\left(\frac{t_1}{\bar{\xi}_2},\frac{t_2}{\bar{\xi}_2}\right)W\left(\frac{t_3}{\bar{\xi}_2}\right) + W^{(2)}\left(\frac{t_1}{\bar{\xi}_2},\frac{t_3}{\bar{\xi}_2}\right)W\left(\frac{t_2}{\bar{\xi}_2}\right) + W^{(2)}\left(\frac{t_2}{\bar{\xi}_2},\frac{t_3}{\bar{\xi}_2}\right)W\left(\frac{t_1}{\bar{\xi}_2}\right) + W\left(\frac{t_1}{\bar{\xi}_2}\right)W\left(\frac{t_2}{\bar{\xi}_2}\right)W\left(\frac{t_3}{\bar{\xi}_2}\right) \tag{45}$$

Generalizing these results, and connecting multi-point factorial correlators with multi-point void probability functions we obtain the result equivalent to eq(29).

$$W^{(s)}\Big(\frac{t_1}{\bar{\xi}_2},\ldots,\frac{t_s}{\bar{\xi}_2}\Big)+\sum_{\rm singlets}W^{(1)}\Big(\frac{t_i}{\bar{\xi}_2}\Big)W^{(s-1)}\Big(\frac{t_1}{\bar{\xi}_2},\ldots,\frac{t_{s-1}}{\bar{\xi}_2}\Big)+\sum_{\rm pairs}W\Big(\frac{t_i}{\bar{\xi}_2}\Big)W\Big(\frac{t_j}{\bar{\xi}_2}\Big)W^{(s-2)}\Big(\frac{t_1}{\bar{\xi}_2},\ldots,\frac{t_{s-2}}{\bar{\xi}_2}\Big) \quad (46)$$
$$+\ldots+\sum_{\rm (s-2)tuplets}W\Big(\frac{t_1}{\bar{\xi}_2}\Big)\ldots W\Big(\frac{t_{s-2}}{\bar{\xi}_2}\Big)W^{(2)}\Big(\frac{t_i}{\bar{\xi}_2},\frac{t_j}{\bar{\xi}_2}\Big)+W\Big(\frac{t_i}{\bar{\xi}_2}\Big)\ldots W\Big(\frac{t_s}{\bar{\xi}_2}\Big)$$
$$=\exp\Big(\sum\frac{\Psi(t_i)}{\bar{\xi}_2}+\sum_{\rm pairs}\frac{\Psi^{(2)}(t_i,t_j)}{\bar{\xi}_2^2}+\sum_{\rm triplets}\frac{\Psi^{(3)}(t_i,t_j,t_k)}{\bar{\xi}_2^3}\ldots+\frac{\Psi^{(p)}(t_1,t_2,\ldots t_k)}{\bar{\xi}_2^p}\Big).$$

It is to be noted that multi-point cumulant correlators can also be defined in such a way that they incorporate all lower order correlations. For example, if we allow one or more of the powers p,q,r or s to vanish the correlator becomes independent of one or more spatial co-ordinates. This will allows a more compact way to express factorial moments in terms of multi-point cumulant correlators (Szapudi & Szalay 1997) which is equivalent to the result obtained using factorial moments. However, it is useful to decompose multi-point factorial moments and multi-point factorial correlators in terms of irreducible one-point and two-point quantities from a computational point of view. As we shall see, this also allows us to relate multi-point cumulant correlators directly with the tree amplitudes involved in the hierarchical ansatz.

The discussion so far is completely general. While Szapudi & Szalay (1993) used the hierarchical approximation of $Q_{N+M}=Q_NQ_M$ and analogous approximations for higher orders, we will be concentrating on the formalism developed by Bernardeau & Schaeffer (1992) in which amplitudes associated with hierarchical trees are assumed to be the product of tree vertices. Following paper I, we can now relate $\Psi^{(n)}(t_1,\ldots,t_n)$ with generating functions $\mu_n(t)$ of vertices appearing in tree-representation of multi-point cumulant correlators which depends on the hierarchical ansatz. These vertices depend on the tree-level amplitudes ν_n in representation of correlation hierarchy of matter correlation function (see paper I for complete discussion):

$$\mu_1(t) = \sum_{n=1}\frac{C_{n1}t^n}{n!} \quad (47)$$
$$\mu_2(t) = \sum_{n=1}\frac{C_{n11}t^n}{n!} \quad (48)$$
$$\mu_3(t) = \sum_{n=1}\frac{C_{n111}t^n}{n!}; \quad (49)$$

etc. The quantities μ_n determine the amplitude associated with a vertex that has n external legs. Using the vertex generating function we can express $\Psi^{(n)}(t_1,\ldots,t_n)$ as:

$$\Psi^{(2)}(t_1,t_2) = \mu_1(-t_1)\xi_{ab}\mu_1(-t_1) \quad (50)$$
$$\Psi^{(3)}(t_1,t_2,t_3)=\mu_1(-t_1)\xi_{ab}\mu_2(-t_1)\xi_{ac}\mu_1(-t_1)+\ldots\text{(cyclic permutations)} \quad (51)$$
$$\Psi^{(3)}(t_1,t_2,t_3,t_4) = \mu_1(-t_1)\xi_{ab}\mu_3(-t_1)\xi_{ac}\mu_1(-t_1)\xi_{ad}\mu_1(-t_1)+\ldots\text{(cyclic permutations)}$$
$$+\mu_1(-t_1)\xi_{ab}\mu_2(-t_1)\xi_{ac}\mu_2(-t_1)\xi_{ad}\mu_1(-t_1)+\ldots\text{(cyclic permutations)} \quad (52)$$

In deriving above relations we have considered only the dominant contribution and this approximation is valid only when the density variance in each cell is much higher than the correlation between cells. Szapudi & Szalay (1997) proposed the use of the two-point cumulant correlator to separate amplitudes associated with different tree vertices, but since the number of tree topologies increases exponentially with order of the diagram (Fry 1984), the two-point cumulant correlator can only be useful in up to 4^{th} order diagrams. This deficiency can only be cured by moving to multi-point cumulants and our formalism developed here will be able to determine ν_n parameters for arbitrary n. However, it is obvious that multi-point cumulants are in general dependent on the geometric configuration of the points, and their determination becomes more complicated as the order increases. We also stress the point that our formalism can in general be used in principle to determine all the tree amplitudes without making any assumptions at all. This is very useful as this can actually test the validity of all hierarchical models of gravitational clustering in the highly nonlinear regime.

The multi-point CPDF (MCPDF) can be decomposed in terms of several one-point quantities. Expanding the multi-point void probability function, which acts as a generating function for MCPDF, Bernardeau & Schaeffer(1992) showed that it is possible to express the MCPDF in terms of a series expansion in $\xi_{ij}/\bar{\xi}_2<<1$, where ξ_{ij} is the two-point correlation function between different overdense cells and $\bar{\xi}_2$ is the variance within one cell as defined above. Such a decomposition is expected to work for overdense cells when they are separated by moderately large distance. To first order in such a decomposition it was found that

$$P(N_1,N_2)=P(N_1)P(N_2)+P(N_1)b(N_1)\xi_{12}(r_{12})P(N_2)b(N_2), \quad (53)$$

where $b(N_1)$ is the bias function for overdense cells which describes how over-dense cells are correlated compared with background mass. Normalisation constrains on $P(N_1,N_2)$ can be translated in to constrains for $b(N)$.

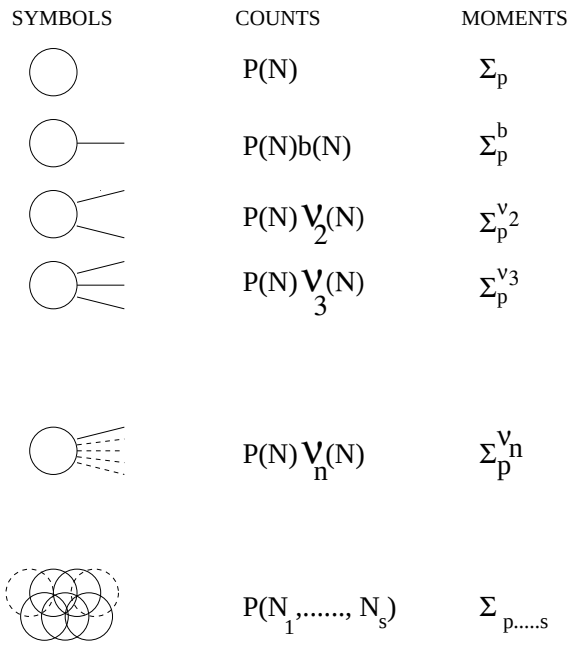

Figure 1. Basic elements in the representation of multi-point joint CPDF and related factorial moments. We represent $P(N)$ and one associated factorial moment Σ_p by a circle with no legs; $P(N)b(N)$ and associated moments Σ_p^b are represented by circle with one leg and, similarly, a circle with n legs represents $P(N)\nu_n(N)$ and its associated factorial moments $\Sigma_p^{\nu_n}$, where p represents the order of the relevant factorial moments, and n the order of the relevant vertex.

$$\begin{aligned} \sum b(N)P(N) &= 0; \\ \sum Nb(N)P(N) &= \bar{N}. \end{aligned} \tag{54}$$

With such a decomposition we can express two-point quantities, such as Σ_{pq}, in terms of one-point factorial moments but with the effect of bias $b(N)$ taken into account:

$$\Sigma_{pq} = \Sigma_p\Sigma_q + \Sigma_p^b\xi_{12}\Sigma_q^b, \tag{55}$$

where we have defined a new biased one point factorial moment Σ_p^b by,

$$\Sigma_p^b = \sum \frac{\xi_2}{p!N_c^p} N(N-1)\dots(N-p+1)b(N)P(N) \tag{56}$$

A similar decomposition can be performed for the 3-point CPDF and, if we write the expression up to second order in $\xi_{ij}\bar{\xi}_2$, we get

$$\begin{aligned} P(N_1,N_2,N_3) &= P(N_1)P(N_2)P(N_3) \\ &+P(N_1)P(N_2)b(N_2)\xi_{23}P(N_3)b(N_3) + P(N_2)P(N_3)b(N_3)\xi_{31}P(N_1)b(N_1) + P(N_3)P(N_1)b(N_1)\xi_{12}P(N_2)b(N_2) \\ &+\nu_2(N_1)P(N_1)\xi_{12}b(N_2)P(N_2)\xi_{13}b(N_3)P(N_3) + \nu_2(N_2)P(N_2)\xi_{23}b(N_3)P(N_3)\xi_{21}b(N_1)P(N_1) \\ &+\nu_2(N_3)P(N_3)\xi_{13}b(N_1)P(N_1)\xi_{32}b(N_2)P(N_2). \end{aligned} \tag{57}$$

A new function appears, together with its associated constraints:

$$\begin{aligned} \sum \nu_2(N)P(N) &= 0 \\ \sum N\nu_2(N)P(N) &= \nu_2\bar{N}; \end{aligned} \tag{58}$$

ν_2 is the second-order vertex associated with a tree-development of the correlation hierarchy in the highly non-linear regime. We define the associated one-point factorial moment Σ_p^ν, which takes into account of corrections second order in $\xi_{ij}\bar{\xi}_2$:

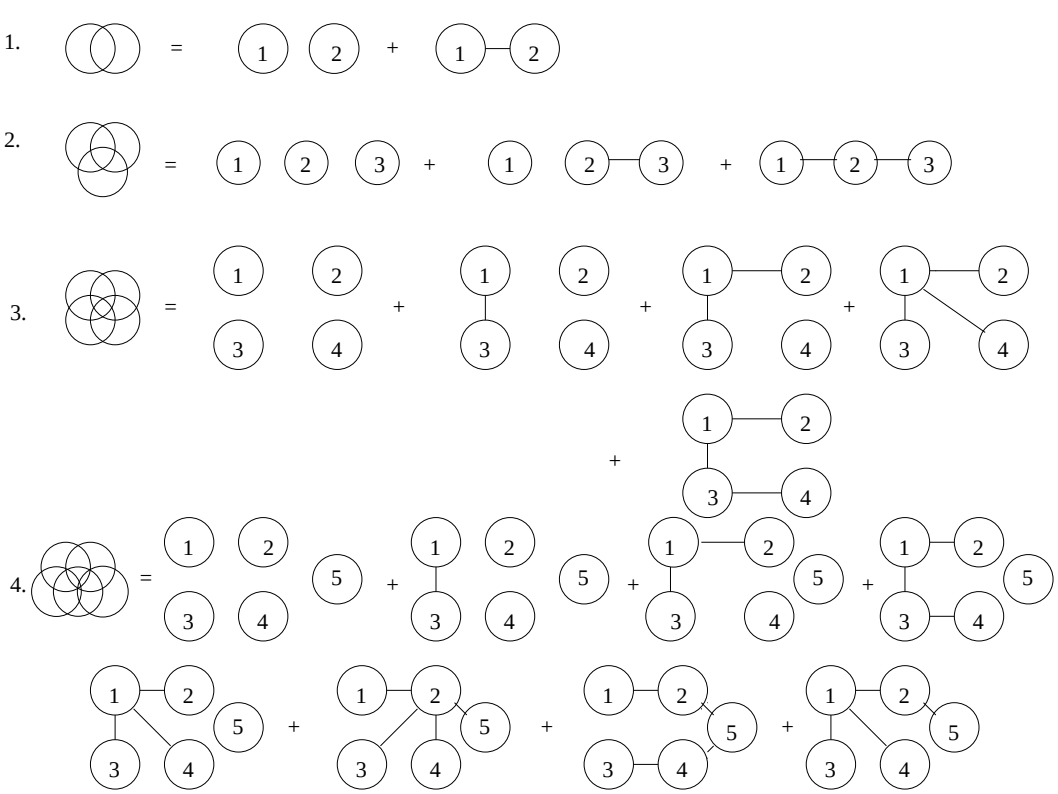

Figure 2. The decomposition of the MCPDF is represented diagrammatically. Every straight line joining two spheres represents the correlation between two cells. Higher-order terms contain more straight lines joining a given set of cells and hence can be ignored in a leading-order approximation. Only tree terms are shown. All possible permutations of labels are to be considered, but in this figure we show only one diagram for each topology. In second order in $\xi_{ij}/\bar{\xi}_i$, it is possible only to have one type of topology, in third order we can have a two different topologies represented by "snake" and "star" diagrams. In even higher order it is possible to have a hybrid of these two different types. It is interesting to note that the correlation hierarchy for matter induces exactly similar hierarchy for MCPDF.

$$\Sigma_p^{\nu_2} = \sum \frac{\xi_2}{p! N_c^p} N(N-1)\dots(N-p+1)\nu_2(N_1)P(N_1). \tag{59}$$

With this higher order contributions are taken into account Σ_{pqr} can be expressed in terms of one point quantities such as Σ_p, Σ_q^b and $\Sigma_{\rm q}^{\nu_2}$,

$$\Sigma_{pqr} = \Sigma_p\Sigma_q\Sigma_r + \Sigma_p\Sigma_q^b\xi_{23}\Sigma_r^b + \Sigma_q\Sigma_p^b\xi_{13}\Sigma_r^b + \Sigma_r\Sigma_p^b\xi_{12}\Sigma_q^b + \Sigma_q^{\nu_2}\xi_{23}\Sigma_p^b\xi_{13}\Sigma_r^b + \Sigma_r^{\nu_2}\xi_{31}\Sigma_p^b\xi_{12}\Sigma_q^b + \Sigma_p^{\nu_2}\xi_{31}\Sigma_p^b\xi_{32}\Sigma_q^b \tag{60}$$

It was shown in paper I that such a decomposition is possible to arbitrary order, as long as we neglect the contributions from loop terms. The tree structure of MPCDF and associated multi-point factorial moments is induced by the inherent tree structure of correlation functions and is very similar in nature. It is interesting to note that, in this particular case, contributions from different orders are completely separated and also the contribution to joint moments from different spatial locations can be expressed as different independent local terms depending only on one spatial co-ordinate. This also makes volume corrections easier to perform, as we shall see in §4.

3 SCALING IN ONE-POINT AND TWO-POINT QUANTITIES

Scaling properties that arise because of the inherent tree structure of correlation functions in highly non-linear regime were studied by Balian & Schaeffer (1989). The principal assumption in this and related studies is that the vertex amplitudes which appear in a tree development of higher-order correlation functions should be constants. Every higher-order correlation function is build by taking suitable products of two-point correlation functions and all such N different tree's carry different amplitudes:

$$\xi_N(\mathbf{r}_1,\dots\mathbf{r}_N) = \sum_{\alpha,\mathrm{N-trees}} T_{N,\alpha} \sum_{\mathrm{labelings}} \prod_{\mathrm{edges}}^{(N-1)} \xi_{(r_i,r_j)}. \tag{61}$$

which in most general case can be expressed as

$$\xi_N(\lambda\mathbf{r}_1,\dots\lambda\mathbf{r}_N) = \lambda^{-\gamma(N-1)}\xi_N(\mathbf{r}_1,\dots\mathbf{r}_N). \tag{62}$$

Clearly no loop terms (such as those that appear in the Kirkwood scaling relation) are considered in such an analysis. There is some support from observations for such an assumption in highly nonlinear regime. On the other hand, in the quasi-linear regime and in the limit of vanishing variance such a hierarchy occurs naturally but the hierarchical coefficients in general depend on different shape parameters appearing in higher-order correlations. Recent studies of the bispectrum, however, show that nonlinearity tends to smooth out such shape dependence and in the highly non-linear regime the bispectrum becomes completely independent of shape. Such a result of course is in agreement with the hierarchical ansatz where all higher-order parameters are assumed to become shape independent (Scoccimarro et al. 1998).

Any assumption regarding the entire hierarchy of many-body correlation functions has very strong consequences and can be used to derive properties of one-point and multi-point statistics. It is important to note that since the exact values of hierarchical coefficients are left unspecified in such an ansatz, it is only possible to talk about generic scaling properties induced by them.

The void probability function (VPF), which is very simple to measure from N-body data or galaxy catalogs carries information on volume averages of all orders of correlation functions. It has been shown that VPF $P_V(0)$ which denotes the probability of finding a cell of volume V empty of any particle can be expressed as an unique function of scaling variable N_c, defined above in §2(White 1979; Balian & Schaeffer 1989). In particular we have (Balian & Schaeffer 1989)

$$P_V(0) = \exp\left(-\bar{N}\sigma(N_c)\right) = \exp\left(-\frac{\phi(N_c)}{\bar{\xi}_2}\right) = \exp\left(\frac{\Psi(-N_c)}{\bar{\xi}_2}\right) \tag{63}$$

For large values of N_c, $\sigma(N_c)$ becomes a power-law with power law index $-\omega$ that depends on initial conditions

$$\sigma(N_c) = aN_c^{-\omega} \tag{64}$$

(Balian Schaeffer (1989). Note that the VPF depends by construction on the discrete nature of sampling and it is not possible to define a VPF for a continuous matter distribution. In effect, however, the VPF can be used to learn about the S_N parameters of the underlying field.

Scaling in VPF induces a similar scaling in the CPDF, and the evolution of the CPDF through different length-scales and different epochs can be described using a unique scaling function $h(x)$ (with $x = N/N_c$). This function is related to the VPF through

$$h(x) = -\int_{-i\infty}^{i\infty} \frac{dy}{2\pi i}\phi(y)\exp(yx) \tag{65}$$

(Balian & Schaeffer 1989) and $P_V(N)$ is related to $h(x)$ by the following expression (Balian & Schaeffer 1989),

$$P_V(N) = \frac{1}{N_c\bar{\xi}_2}h\left(\frac{N}{N_c}\right) \tag{66}$$

Using these relations it is now possible to express $P_V(N)$ as

$$P_V(N) = \frac{a}{\bar{\xi}_2 N_c}\frac{1-\omega}{\Gamma(\omega)}\left(\frac{N}{N_c}\right)^{\omega-2} \tag{67}$$

(Balian & Schaeffer 1989). This expression is valid in the range $N_v < N < N_c$ where, physically, N_v is the typical cell occupancy in underdense regions and N_c in overdense regions; these are related to each other by $N_v = N_c(a/\xi_2)^{1/(1-\omega)}$. It is interesting to note that the power-law index $-\omega$, which appears in the expression of the VPF, also appears in case of CPDF.

For highly overdense cells a different asymptote can be computed by noting the that $\phi(y)$ exhibits a singularity for small negative values of $y = y_s$

$$\phi(y) = \phi_s - a_s\,\Gamma(\omega_s)\,(y-y_s)^{-\omega_s} \tag{68}$$

It introduces an exponential cut-off for large values of N in $P(N)$ (Balian & Schaeffer, 1989),

$$h(x) = a_s\,x^{\omega_s-2}\exp(-|y_s|\,x), \tag{69}$$

Previous studies have shown that an analytical fitting function for $h(x)$ works well in both these regimes (Balian & Schaeffer 1989):

$$h(x) = \frac{a(1-\omega)}{\Gamma(\omega)}\frac{x^{\omega-2}\exp(-x|y_s|)}{(1+bx)^c} \tag{70}$$

The constraints satisfied by $h(x)$ are

$$\int_0^\infty xh(x)dx = 1 \tag{71}$$

and and

$$\int_0^\infty x^2h(x)dx = 1. \tag{72}$$

The hierarchical ansatz has also been used to predict a scaling behaviour of the multi-point CPDF and related statistics. The formalism developed by Bernardeau & Schaeffer (1992) shows that the two-point CPDF can be written as

$$P_V(N_1, N_2) = P(N_1)P(N_2)\left[1 + \xi_2(r_1, r_2)b(N_1)b(N_2)\right], \tag{73}$$

where we have denoted joint occupation probability of two different cells (we take sizes of these two cells to be the same) by $P_V(N_1, N_2)$. Such a decomposition means that the correlation function of two cells with occupancy N_1 and N_2 (Bernardeau & Schaeffer 1992) is

$$\xi_2(N_1, N_2) = b(N_1)b(N_2)\xi_2(r_1, r_2). \tag{74}$$

Several characteristic features of the bias induced by these scaling properties can be noted. For one thing, it is clear from this derivation that the bias depends only on x and hence only on intrinsic properties of the collapsed object, at least for large separations. Since the bias $b(N)$ satisfies the property $\sum_N P(N)b(N) = 0$ it is clear that if the bias for some occupation numbers is positive, it must be negative for others. Detailed calculations show that cells with occupation number close to N_v (the average occupation of cells in underdense regions) have negative bias. On the other hand, the correlation of two identical objects (i.e. cells with same occupation number) is always positive:

$$b(x) = \left(\frac{\omega}{2a}\right)^{1/2} \frac{\Gamma[\omega]}{\Gamma[\frac{1}{2}(1+\omega)]} x^{(1-\omega)/2}; \qquad (\text{for} \quad N_v < N < N_c) \tag{75}$$

$$b(x) = \frac{x}{x_{bs}}; \qquad (\text{for} \quad N > N_c) \tag{76}$$

Interesting though $b(x)$ is in itself, from a computational point of view it is much easier to compute $b(> x)$, the bias of overdense cells above certain threshold, defined as

$$b(> x) = \frac{\int_x^\infty b(x')h(x')dx'}{\int_x^\infty h(x')dx'}. \tag{77}$$

It can be shown in the two regimes we have discussed that $b(> x)$ can be related directly with $b(x)$, i.e. $b(> x) = 2b(x)$ for $N < N_c$ and $b(> x) = b(x)$ for $N > N_c$. It is interesting to point out at this point that for a Gaussian density field $b(> x) = b(x) = 1$ for rare events. In the quasilinear regime, the value of $b(> x)$ computed numerically has been shown to match theoretical predictions (Bernardeau 1996). However no such studies have been made in the highly nonlinear regime where we expect the hierarchical ansatz to be valid.

In paper I we derived higher-order predictions of the hierarchical ansatz, but we restrict ourselves here to two-point quantities associated with overdense cells. The bias in this case acts as a generating function for the cumulant correlators; this was studied in Munshi & Melott (1998). A complete test of higher-order predictions covering the CPDF and VPF of overdense cells will be presented elsewhere.

4 SIMULATIONS AND DATA ANALYSIS

4.1 Simulations

The simulations used here are numerical models for the gravitational clustering of collisionless particles in an expanding background. We study evolution of initial two-dimensional Gaussian perturbations in $\Omega = 1$ universe. All the 2D simulations are done with a particle-mesh (PM) code with 4096^2 particles with an equal number of grid points. The code has at least twice the dynamical resolution of any other PM code with which it has been compared. The 2D simulations we use here are very similar to those described in detail in Beacom et. al (1991), a video of their evolution in shown in Kauffmann & Melott (1992), but with 64 times more particles. In this paper we analyze a subset of the simulations with featureless power-law initial spectra of the general form,

$$\begin{aligned} P(k) &\propto k^n \text{ for } k \le k_c, \\ &= 0 \text{ for } k > k_c. \end{aligned} \tag{78}$$

We have analysed power-law models with power spectral index $n = 2, 1, 0, -1$ in 2D, with a cutoff in each case at the Nyquist wave number $k_c = 2048\, k_f$, where $k_f = 2\pi/L_{\rm box}$ is the fundamental mode associated with the box size.

We choose $\sigma(k_{\rm NL})$, the epoch when the scale $2\pi/k_{\rm NL}$ is going nonlinear, as a measure of time. The first scale to go nonlinear is the one corresponding to the Nyquist wave number. This happens, by definition, when the variance σ is unity. As σ increases, successive larger scales enter in the nonlinear regime. The simulations were stopped at $\lambda_{\rm NL} = 2l_{\rm grid},\ 4l_{\rm grid},\ 8l_{\rm grid}, ...,\ L_{\rm box}/2$. In our study we have studied epochs till $L_{\rm box}/16$ goes nonlinear.

The construction of $h(x)$ uses measurements of the CPDF from several different epochs. We concentrated on smaller scales which are already in the highly nonlinear regime. Cell sizes of $(4l_{\rm grid})^2$, $(8l_{\rm grid})^2$ and $(16l_{\rm grid})^2$ were studied for four different epochs when $L_{\rm box}/8$, $L_{\rm box}/16$, $L_{\rm box}/32$ and $L_{\rm box}/64$ have gone nonlinear.

The growth rates of various modes in the linear regime were studied by Melott et al. (1988) for this PM code. The results at $\lambda = 3l_{\rm grid}$ are equivalent to those obtained by a typical PM code at $\lambda = 8l_{\rm grid}$, owing to the staggered mesh scheme.

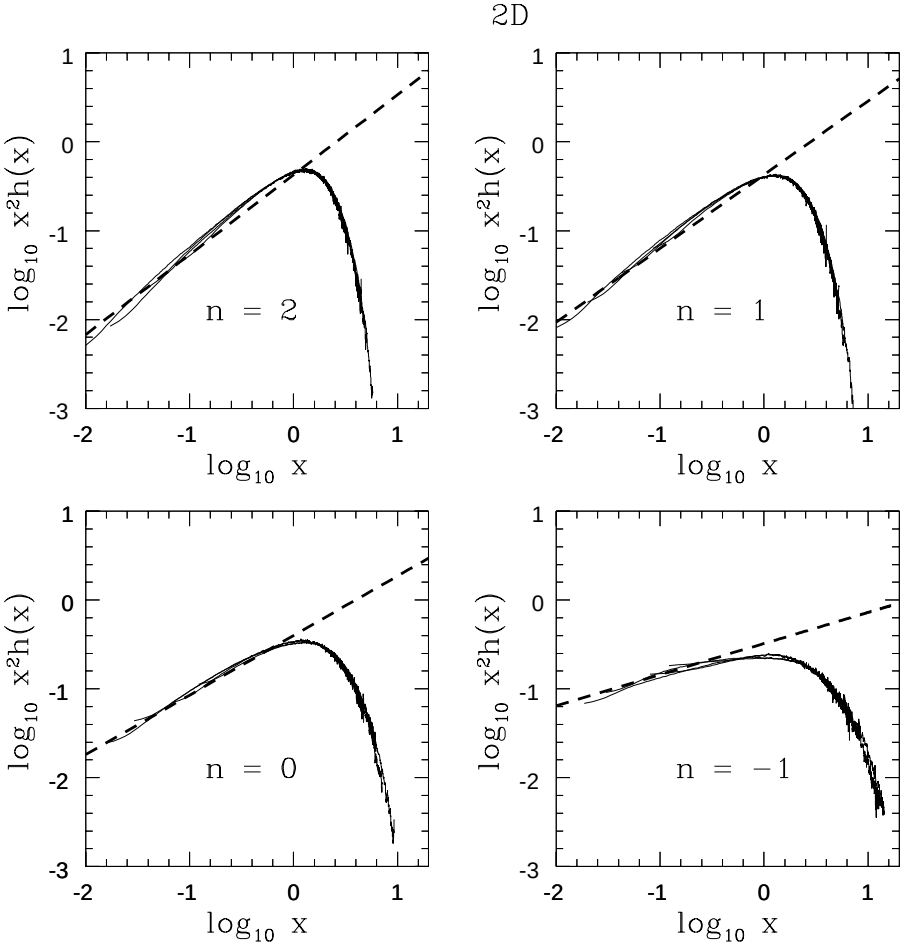

Figure 3. The measured $h(x)$ in 2D numerical simulations, as described in the text. The large dynamic range allows us to determine the function $h(x)$ to an accuracy at the level of 10^{-3}, and order-of-magnitude increase compared to all previous studies in 2D or 3D. The function $h(x)$ in general shows a power law profile for $x < 1$ which is followed by an exponential cutoff for $x > 1$. This exponential cutoff is more severe when there are more power at smaller scales. Since $x = N/N_c$ and N_c is the typical occupation of cells inside clusters, this means that the CPDF $P_V(N)$ at any scale will have a power-law profile for $N < N_c$ and will have an exponential tail for $N > N_c$. In an infinite catalogue this exponential tail will extend to infinitely large values of N. However, in a finite catalogue, $P_l(N)$ shows an abrupt cutoff at $N_{\rm max}$, due to the absence of any denser cells beyond that limit. Different length scales map different regions of $h(x)$ curve. Previous studies have shown that effect of finite volume corrections increases with power on larger scales which reduces the available range of x for which $h(x)$ can be measured accurately. The measured $h(x)$ satisfies two integral constraints ($S_1 = S_2 = 1$) and the power law index of $x^2h(x)$ should be exactly the same as the negative power-law index appearing in $\sigma(N_c)$.

So we expect that our code performs well at the wavelength associated with four cells and since the collapse of $4\,l_{\rm grid}$-size perturbations will give rise to condensations of diameter $2\,l_{\rm grid}$ or less, the smallest cell size that can be safely resolved is $2\,l_{\rm grid}$. The epochs we include in our study are not affected by the fact that simulations were started by Zel'dovich approximation.

4.2 Data Analysis

For analysing the data, computations of the count probability distribution function (CPDF) $P_l(N)$ and 2CPDF $P_{l,r}(N_1, N_2)$ were performed by laying down a grid of mesh spacing l and counting the occupation number in each cell; this yields the probability of finding N particles in cell size l and also the joint probability of finding N_1 and N_2 particles in two different cells separated by a distance r. Statistics were improved by perturbing the grid in each orthogonal direction and keeping the mesh undistorted while repeating the counting process. We considered cells of size $4l_{\rm grid}$, $8l_{\rm grid}$ and $16l_{\rm grid}$ in our studies. With these cell sizes we can reach probabilities as small as few times 10^{-9}. We have also checked that our results do not change if we increase the number of sampling cells; we get exactly the same results even with ten times fewer cells. For computing the VPF we have only considered cell sizes where the condition $P_0 > 1/e$ is satisfied; it was shown by Colombi et al. (1995) that the properties of voids larger than this are affected by the grid which is used to start simulations.

To extend the range of N_c for which we can study the scaling properties we have diluted the data while studying scaling properties of VPF. Different levels of dilution such as 2048^2, 1024^2, 512^2 and 256^2 were considered. Results of the VPF

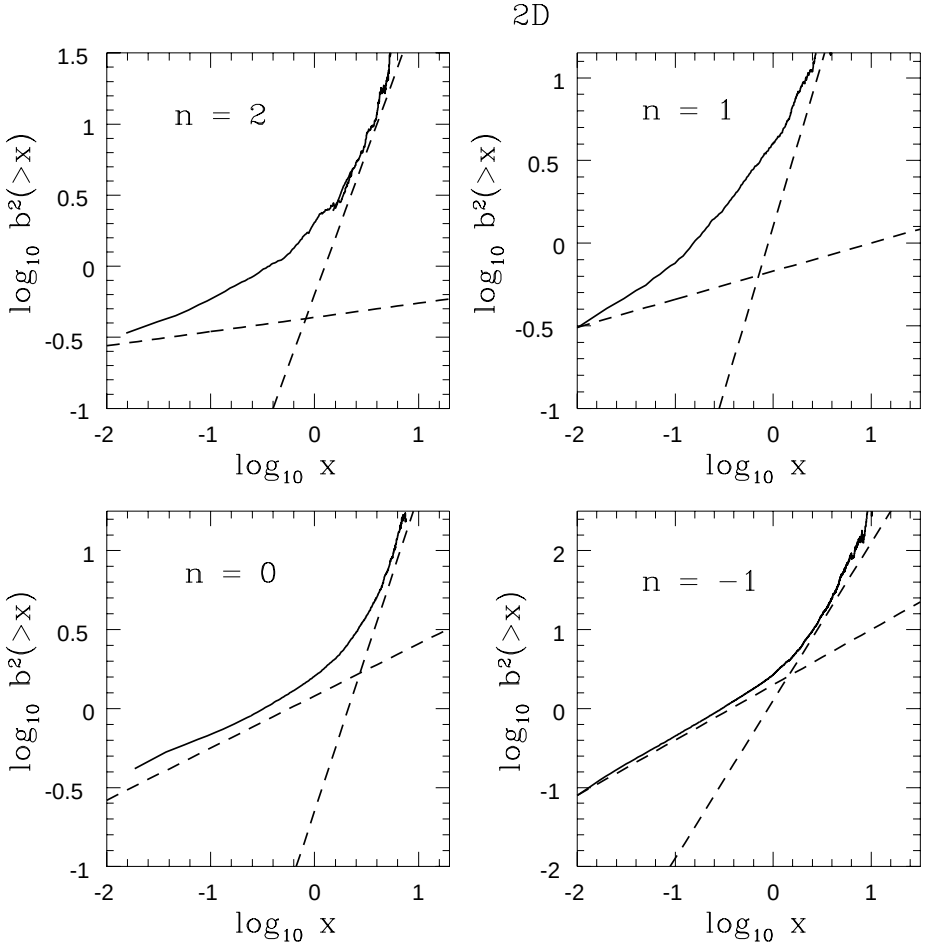

Figure 4. The measured values of $b(> x)$ from the 2D simulations described in the text. The bias $b(> x)$ shows two different regimes similar to $h(x)$: for small values of x, it increases slowly with x according to a power-law with index less than unity. This index is also related to the negative slope of $\sigma(N_c)$ and slope of $x^2 h(x)$ for $x < 1$. For larger values of x, $b(> x)$ shows a steeper increase, proportional to x. For $b(N)$ this means that moderately overdense cells show a slow increase of bias with cell occupancy N but for highly overdense cells the bias increases linearly. The transition occurs for $N = N_c$ which is the typical occupancy of cells in over-dense regions. We have used two related but different methods to estimate bias. The first method is based on measuring cross-correlation of different classes of objects with respect to background mass distribution and the second method is based on computation of direct measurement of correlation function against the background mass distribution for different classes of objects. Good agreement between these two methods proves that the factorization property of bias and also shows that the bias for an overdense cell depends only on the intrinsic property of the object concerned and can be expressed as a unique function of x.

analysis are plotted in Figure 5, showing that the scaling ansatz predicts the VPF very well. We find a power-law profile for $\sigma(N_c)$ for large values of N_c. At small values of N_c, when the data is dominated by Poisson noise, $\sigma(N_c)$ tends to unity. Our results of analysis of VPF proves beyond doubt that the hierarchical ansatz is indeed a good approximation in the highly non-linear regime. Since $\sigma(N_c)$ acts as a generating function for the S_N parameters this means that all S_N parameters reach a constant value in this limit. Furthermore, as the S_N parameters are linear combinations of product of powers of amplitudes of different vertices that appear in the tree-level correlation hierarchy, this will also mean that these vertices too are constant in highly non-linear regime.

However, we have also found a significant departure from such scaling for $n = -2$ spectra. The maximum value of N_c for which the VPF can be probed reliably decreases with increasing power on larger scales and the simulation grid, which remains undistorted in underdense regions for spectra with more power at larger scales even in the highly non-linear regime. these two factors may be responsible for a spurious departure from scaling in this case. Strong motivation for suspecting that the apparent departure we detect in scaling for VPF for $n = -2$ may not be real is that we do not detect any corresponding effect in $h(x)$.

We use four neighbouring equidistant cells in a grid to evaluate the bias with respect to the central cell. We adopt two different methods to evaluate the bias which, as suggested by Bernardeau (1994), test the factorizability property of bias of over-dense cells predicted by the hierarchical ansatz. In the first method we compute the bias associated with each overdense cells with respect to background mass. This is done by casting the definition of bias in a following way:

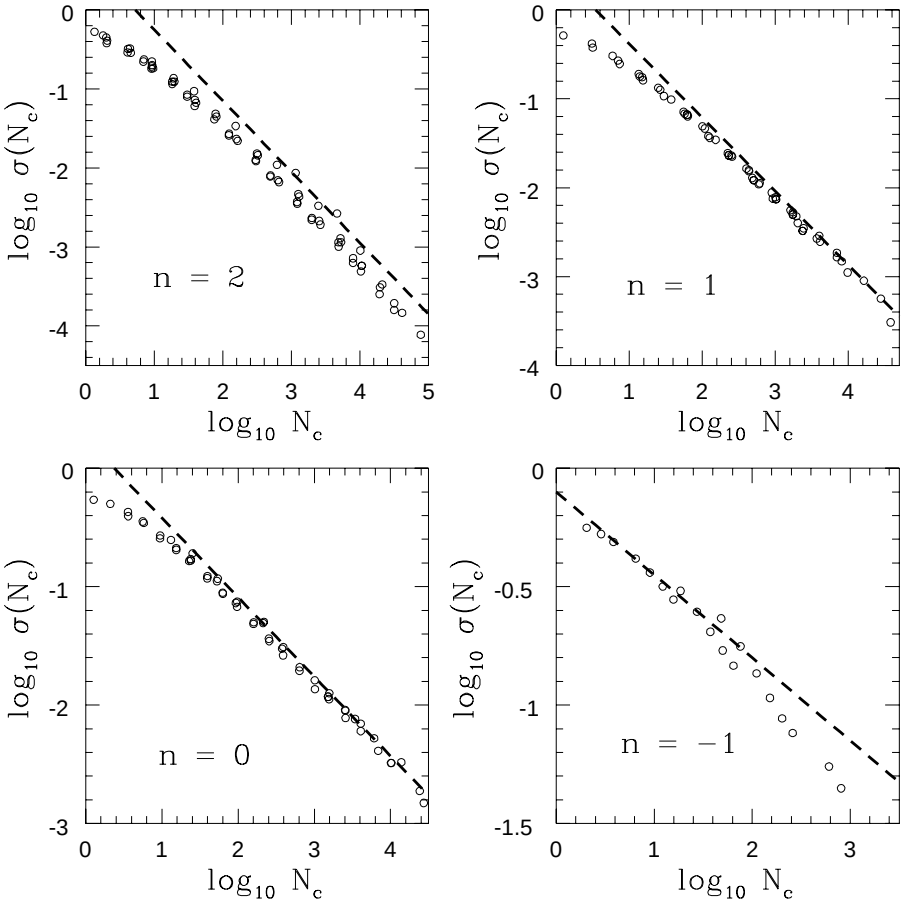

Figure 5. The measured values of $\sigma(N_c)$ from 2D simulations described in the text. For a given cell size, the parameter N_c increases as the system evolves and for a given epoch, the larger cells have higher values of N_c. Cell size we analyze ranges from $(4l_{\rm grid})^2$ to $(64l_{\rm grid})^2$. We have used different epochs of nonlinearity to construct the function $\sigma(N_c)$. We also use different levels of nonlinearity to increase available range of N_c. Different dilution levels of the 4096^2 data, i.e. 2048^2, 1024^2, 512^2 and 256^2 were used to test the validity of the scaling ansatz. For power spectra $n = 2, 1$ and 0 we do not detect any deviation from scaling models, but for $n = -2$ there are discernible departures. This could be due to various spurious effects which could influence the determination of $\sigma(N_c)$, especially the initial grid which survives undistorted in underdense regions if the initial spectrum is $n = -2$. Error estimation in determination of $\sigma(N_c)$ was done by Colombi et al. (1995) who found that the error is proportional to S_4 which increases with more power on larger scales. However all our data points satisfy the criteria $P_0 > 1/e$ for avoiding errors introduced by the computational grid. This restricts the maximum cell size for a given epoch for which $\sigma(N_c)$ can be computed. The other possibility for violation of particular scaling model which we have considered here arises from the fact that instead of S_N parameters being constant at highly nonlinear regime they might show a very slow but steady increase with level of nonlinearity as reported by Lucchin et. al (1994) and Colombi et. al (1996). However note that such an argument will also mean a clear departure of CPDF from scaling which we have not detected in our studies.

$$b(>N) = \Big(\frac{\sum_{N_1>N} P(N_1)b(N_1)}{\sum_{N_1>N} P(N_1)}\Big) = \frac{1}{\xi_{ij}}\Big(\frac{\frac{1}{N_2}\sum_{N_1>N}\sum_{N_2} N_2 P(N_1,N_2)}{\sum_{N_1>N} P(N_1)} - 1\Big). \tag{79}$$

In the second method we directly compute the bias from correlation of two different over-dense cells:

$$b(>N) = \Big(\frac{\sum_{N_1>N} P(N_1)b(N_1)}{\sum_{N_1>N} P(N_1)}\Big) = \sqrt{\Big(\frac{1}{\xi_{ij}}\Big(\frac{\sum_{N_1>N}\sum_{N_2>N} P(N_1,N_2)}{(\sum_{N_1>N} P(N_1))^2} - 1\Big)\Big)} \tag{80}$$

In both cases we compute the bias for cells whose occupancy is greater than some particular threshold x and then we vary the threshold to study $b(>x)$ for different values of x. We initially attempted to compute directly the function $b(x)$ but we found measurements of the cumulative distribution $b(>x)$ to be much more stable. We also found that the bias computed from both methods mentioned above matches, and also the slope of the bias function for overdense cells agrees well with theoretical predictions. In particular, the bias increases with x as a power law with the index of power-law being less than unity, whereas for highly overdense cells the bias is directly proportional to x. Determination of the bias is more difficult for spectra with more power on smaller scales, due to absence of long range correlations in such models.

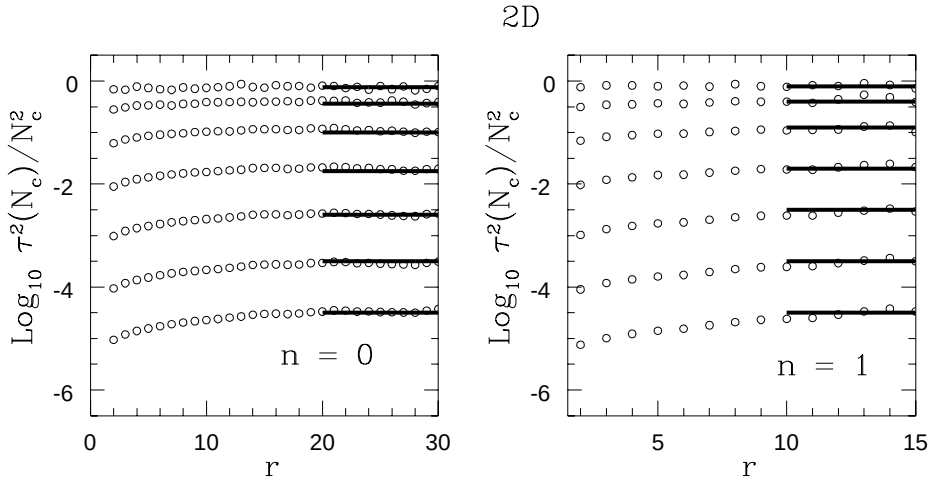

Figure 6. We plot the quantity $\tau^2(N_c)/N_c^2$ described in the text for different cell separations r. We have shown results for only two different initial conditions for one particular cell size of $(4l_{\rm grid})^2$. For each cell size we consider different levels of dilutions. The lowest curve in each panel correspond to 4096^2 data. Each subsequent curve from bottom to top correspond to a dilution by a factor of 1/4 in $\bar{N}$, while the top-most curve correspond to 64^2 data. The solid lines at larger values of r correspond to values of $\tau^2(N_c)/N_c^2$ taken for scaling studies; this quantity exhibits scaling properties when plotted as a function of N_c as shown in Figure 7. Departure from linear behaviour for small separation is due to terms higher order in $\xi_{ij}/\bar{\xi}_i$ which are neglected in our study of the 2VPF.

Table 1. Parameters of the fitting function $h(x)$ and $b(x)$ in 2D

	-1	0	1	2
ω	.35	.67	.85	.9
a	1.25	1.778	2.81	4.46

Computation of the 2VPF was done using a similar technique as in the computation of bias. We used four neighbouring cells for every cell to find the joint probability of finding the pair of cells to be devoid of any particles; we did this for several different separations of two cells. For a given value of the scaling parameter N_c, we have also studied the scaling properties associated with the 2VPF for larger separation. Data was diluted to from 4096^2 to 2048^2, 1024^2, 512^2 and 256^2 to increase the available dynamic range of the scaling parameter N_c as we did for the case of VPF. For large cell separations, where all higher order terms in $\xi_{12}/\bar{\xi}$ are negligible,

$$\frac{1}{\xi_{12}}\frac{1}{\bar{N}^2}\log\left(\frac{\mathrm{P}(0,0)}{\mathrm{P}(0)^2}\right) = \frac{\tau^2(\mathrm{N_c})}{\mathrm{N_c^2}}. \tag{81}$$

We found that the scaling function $\tau(N_c)^2/N_c^2$, which encodes the scaling properties of the 2VPF and also acts as a generating function for cumulant correlator (Munshi & Melott 1998) is described very accurately by the theoretical prediction, made on the basis of the hierarchical ansatz; see Figures 6 & 7.

5 FINITE VOLUME CORRECTION

The finite size of galaxy catalogues and numerical simulations always affects the determination of many-body correlation functions from real or simulated samples. The higher the order of correlation measured, the greater the contribution to it from exceptionally dense and consequently rare clumps. Their determination is therefore always affected by the presence or absence of such clumps in a finite sample. This manifests itself in a sharp cutoff in all the one-point functions we have defined such as $P(N)$, $b(N)$, and $\nu_2(N)$. Using the scaling arguments we have already described n it is however possible to supplement the measured quantities with the functions by $h(x)$, $b(x)$, and $\nu_2(x)$. These quantities have been computed by series expansion of multi-point VPF in increasing power of $(\xi_{ij}/\bar{\xi}_i)$ for second and third order by Bernardeau & Schaeffer (1992) and for fourth and fifth order in our Paper I. Using their results we show here that scaling arguments once verified at lower order in $(\xi_{ij}/\bar{\xi}_i)$ can be used to estimate finite volume errors for in estimates of the multi-point cumulant correlators.

If $N_{\rm max}$ is the maximum occupancy of cells of certain length scales in a catalog, measured $P(N)$ for $N < N_{\rm max}$ shows an abrupt cutoff for $N = N_{\rm max}$. We supplement the measured $P(N)$ by $\Pi(\nu)$ for $N > N_{\rm max}$ to construct the corrected $P^{\rm c}(N)$

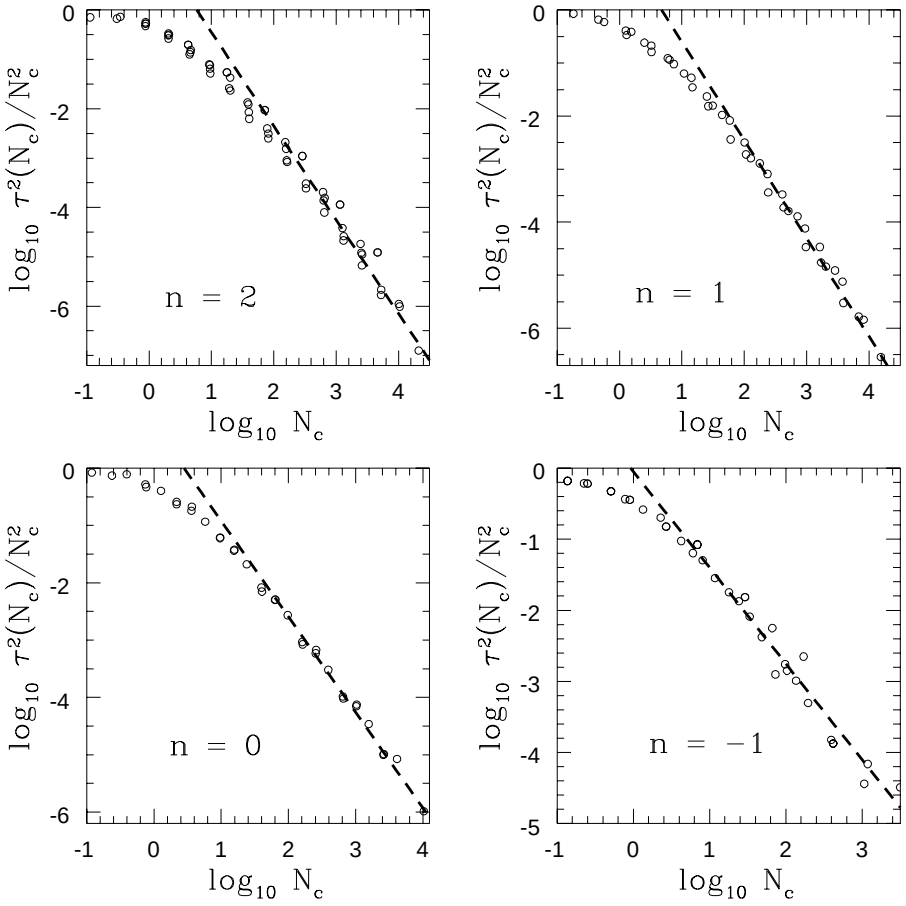

Figure 7. The measured values of $\tau^2(N_c)/N_c^2$ from 2D simulations described in the text. We compute the 2VPF $P(0,0)$ and VPF $P(0)$ using a regular grid for different separation as explained in previous figures and use this to estimate $\tau^2(N_c)/N_c^2$ for different values of N_c in the limit of large separations as described in the text. As in the case of VPF we repeat the process for different values of cell size from $(4l_{\rm grid})^2$ to $(64l_{\rm grid})^2$. We also use different levels of dilution to increase the range of N_c studied. We study 2VPF for 4096^2, 2048^2, 1024^2, 512^2, 256^2, 128^2 and 64^2 data points. All length scales which do not satisfy the criteria $P_0 < 1/e$ were excluded from analysis to avoid lattice effects. We also study different levels of nonlinearity and cell sizes as in the analysis of VPF. Dashed lines correspond to the theoretical prediction $\tau^2/y^2 \propto y^{-(1+\omega)}$, where $-\omega$ is also the slope of one point VPF.

(Munshi et al. 1998c); $\Pi(\nu)$ can be expressed in terms of the scaling function $h(x)$ describing $P(N)$:

$$P^{\rm c}(N) = \Big(1 - \frac{\mathcal{A}}{\bar{\xi}_2} - \mathcal{B}\frac{(N-\bar{N})}{N_c}\Big)P(N) \qquad N < N_{\rm max} \tag{82}$$
$$P^{\rm c}(N) = \Pi(\nu) \qquad N > N_{\rm max}.$$

The corrected CPDF satisfies two constraint equations similar to the case of the uncorrected CPDF. The constraints

$$\sum P(N) = 1, \quad \sum P^{\rm c}(N) = 1, \quad \sum NP(N) = \bar{N}, \quad \sum NP^{\rm c}(N) = \bar{N} \tag{83}$$

can be used to determine the constants $\mathcal{A}$ and $\mathcal{B}$ where

$$\mathcal{A} = H_0 = \int h(x)dx \tag{84}$$

$$\mathcal{B} = H_1 = \int xh(x)dx \tag{85}$$

and

$$\Sigma_p{}^{\rm c} = \Sigma_p - (p+1)H_1\Sigma_{p+1}^{\nu_s} + \frac{H_p}{p!}. \tag{86}$$

Similarly, the bias associated with overdense cells $b(N)$ shows an abrupt cutoff at $N > N_{\rm max}$ which we supplement by its

continuous analogue $b(\nu)$, which can be expressed in terms of its scaling function $b(x)$:

$$\begin{aligned} P^{\rm c}(N)b^{\rm c}(N) &= \Big(1-\mathcal{C}-\mathcal{D}\big(N-\frac{\langle N^2\rangle_b}{\bar{N}}\big)\Big)P(N)b(N) \qquad N<N_{\rm max} \\ P^{\rm c}(N)b^{\rm c}(N) &= \Pi(\nu)b(\nu) \qquad\qquad\qquad\qquad\qquad\quad N>N_{\rm max}, \end{aligned} \tag{87}$$

where we have denoted $\sum N^2P(N)b(N)$ by $\langle N^2\rangle_b$. The unknown parameters $\mathcal{C}$ and $\mathcal{D}$ can be determined by using the fact that both $b(N)P(N)$ and corrected $b^{\rm c}(N)P^{\rm c}(N)$ satisfy similar constraints. For the zeroth moment: $\sum b(N)P(N)=0$, $\sum b^{\rm c}(N)P^{\rm corr}(N)=0$ and similarly for the first moment we have $\sum Nb(N)P(N)=\bar{N}$, $\sum Nb^{\rm c}(N)P^{\rm c}(N)=\bar{N}$. The constraint on the zeroth-order moment can be used to determine $\mathcal{D}$, and the first-order constraint fixes the value of $\mathcal{C}$:

$$\mathcal{C} = B_1 = \int_{N_{\rm max}/N_c}^{\infty} xh(x)b(x)dx \tag{88}$$

$$\mathcal{D} = B_0 = \int_{N_{\rm max}/N_c}^{\infty} h(x)b(x)dx \tag{89}$$

and

$$\Sigma_p^{b\,{\rm c}} = (1-B_1)\Sigma_p^b - (p+1)B_0\Sigma_{p+1}^b + \frac{B_p}{p!}. \tag{90}$$

Where we have defined the p^{th} order moment of $b(x)h(x)$ by B_p and

$$B_p = \int_{N_{\rm max}/N_c}^{\infty} x^p b(x)h(x)dx. \tag{91}$$

Extending these calculations to the case of MCPDFs is fairly straightforward. The constraints $\sum P(N)\nu_s(N)=\bar{N}\nu_s$, $\sum NP^{\rm c}(N)\nu_s^{\rm c}(N)=\bar{N}\nu_s$ and $\sum P(N)\nu_s(N)=1$, $\sum P^{\rm c}(N)\nu_s^{\rm c}(N)=1$ can be used to determine the values of $\mathcal{X}$ and $\mathcal{Y}$ appearing in the renormalization of $P(N)\nu_s(N)$ that corrects for the finite volume effect:

$$\begin{aligned} P^{\rm c}(N)\nu_n^{\rm c}(N) &= \Big(1-\mathcal{X}-\mathcal{Y}\big(N-\frac{\langle N^2\rangle_{\nu_n}}{\bar{N}}\big)\Big)P(N)\nu_n(N) \qquad N<N_{\rm max} \\ P^{\rm c}(N)\nu_n^{\rm c}(N) &= \Pi(\nu)\nu_n(\nu) \qquad\qquad\qquad\qquad\qquad\quad N>N_{\rm max}, \end{aligned} \tag{92}$$

where $\mathcal{Y}$ and $\mathcal{X}$ are the first and zeroth order moments of $h(x)\nu_n(x)$ respectively and $\langle N^2\rangle_{\nu_n}$ represents $\sum N^2P(N)\nu_n(N)$; these quantities can be expressed by

$$\mathcal{X} = Z_1^{\nu_n} = \frac{1}{\nu_n}\int_{N_{max}/N_c}^{\infty} xh(x)\nu_n(x)dx \tag{93}$$

$$\mathcal{Y} = Z_0^{\nu_n} = \frac{1}{\nu_n}\int_{N_{\rm max}/N_c}^{\infty} h(x)\nu_n(x)dx \tag{94}$$

and

$$\Sigma_p^{\nu_n\,{\rm c}} = (1-Z_1^{\nu_n})\Sigma_p^{\nu_n} - Z_0^{\nu_n}(p+1)\Sigma_{p+1}^{\nu_n} + \frac{Z_p}{p!}, \tag{95}$$

where we have denoted the p^{th} moment of $\nu_n(x)h(x)$ by

$$Z_p^{\nu_n} = \frac{1}{\nu_n}\int_{N_{max}/N_c}^{\infty} x^p h(x)\nu_n(x)dx. \tag{96}$$

If we also notice that we have to correct N_c for the finite volume effect using $N_c^{\rm c}=N_c(1-6H_1\Sigma_3+H_2)$. Incorporating this correction we can finally write

$$\Sigma_p{}^{\rm c} = \frac{\Sigma_p-(p+1)H_1\Sigma_{p+1}^{\nu_s}+\frac{H_p}{p!}}{(1-6H_1\Sigma_3+H_2)^{p-1}} \tag{97}$$

$$\Sigma_p^{b\,{\rm c}} = \frac{(1-B_1)\Sigma_p^b-(p+1)B_0\Sigma_{p+1}^b+\frac{B_p}{p!}}{(1-6H_1\Sigma_3+H_2)^{p-1}} \tag{98}$$

$$\Sigma_p^{\nu_n\,{\rm c}} = \frac{(1-Z_1^{\nu_s})\Sigma_p^{\nu_n}-Z_0^{\nu_n}(p+1)\Sigma_{p+1}^{\nu_n}+\frac{Z_p^{\nu_s}}{p!}}{(1-6H_1\Sigma_3+H_2)^{p-1}}. \tag{99}$$

We have kept only terms which are leading order in $1/\bar{\xi}$ in the expressions and hence they are valid only in the highly nonlinear regime. We have also ignored a term in $\langle N^2\rangle_{\nu_s}/\bar{\xi}_2\nu_n\bar{N}^2$ since we expect $\langle N^2\rangle_{\nu_n}/\nu_n\bar{N}^2$ to be of order unity. The inclusion of terms representing loops of lower-order diagrams in the calculations will bring not only moments of $b(x)h(x)$ and $\nu_n(x)h(x)$ but also moments of powers of these scaling functions which we have neglected here.

Cumulants and cumulant correlators have been measured with presently available N-body and galaxy catalogues. Although

several analysis for finite volume correction for cumulants are available in the literature (Colombi et al. 1992, 1994, 1995, 1996, Munshi et al. 1998), no such analysis has been done for cumulant correlators. Our analysis based on the hierarchical ansatz provides an estimate for finite volume correction and also can be used to correct such effects. Implementation of our method for measurements of MCCs will be presented elsewhere.

It was shown by Szapudi & Szalay (1997) that cumulant correlators can actually be used to separate amplitudes associated with different tree topologies at low order. However, at higher orders the number of equations becomes less than the number of independent tree topologies rendering the system indeterminate. This can only be cured if we increase the number of points by using multi-point cumulant correlators. This in principle will allow us not only to determine the amplitudes associated with tree topologies of arbitrary order but also allow an additional consistency check because the number of variables is less than the number of equations they have to satisfy. Of course going from two-point cumulant correlators to multi-point cumulant correlators makes numerical computation more complicated, especially because of the shape dependence of multi-point cumulant correlators for more than two points. With the availability of simulations with much larger dynamic range a determination of the ν_n parameters will become possible in the reasonably near future.

The procedure which we have introduced here for finite volume corrections however requires *a priori* knowledge of the ν_n parameters which one needs to determine from the same sample. This calls for an iterative procedure which can be applied until the whole process becomes convergent.

6 CONCLUSIONS

In this and the previous paper (Paper I), we have investigated properties of the multi-point cumulant correlators - the natural generalisations of the one-point cumulants and the two-point cumulant correlators. We also developed a method which can be used to determine the amplitudes associated with tree diagrams of different topologies, without making any simplistic assumptions about them. This method can be used to test any models of gravitational clustering in the highly non-linear regime, such as the hierarchical scaling ansatz.

We have developed a diagrammatic method to represent the decomposition of multi-point cumulant correlators in terms of factorial moments and factorial correlators. Using this approach, we have found that every higher-order contribution to $(\xi_{ij}/\bar{\xi}_i)$ introduces a new function which corresponds to a new vertex in the tree-level hierarchy, labeled by $\nu_n(N)$. It was shown in Paper I that these functions display similar scaling relations as the functions $P(N)$ and $b(N)$. We have used these scaling relations for $\nu_n(N)$ to estimate the effect of the finite size of N-body simulations or galaxy catalogues in estimates of statistical descriptors, such as cumulants, cumulant correlators and multi-point cumulant correlators.

We have tested various predictions of the hierarchical ansatz, relating to one-point and two-point quantities in the highly non-linear regime, using high-resolution 2D numerical simulations. We find that the scaling functions $h(x)$ and $\sigma(N_c)$ can describe the behaviour of the CPDF and VPF in highly non-linear regime for all length scales, as the hierarchical assumption predicts. These functions were, however, found to depend on initial conditions even in the highly nonlinear regime. As expected from general scaling arguments, it was found that $\sigma(N_c)$ attains a asymptotic power-law $N_c^{-\omega}$ for large values of N_c. However, such models are not able to predict the values of ω; more detailed dynamical arguments would have to be developed to make such a prediction. Numerically, however, we have found that ω depends on the form of the initial power spectrum, increasing with the relative amount of small-scale power. The index ω also enters in the description of $x^2h(x)$ which we found to have a power law profile with power law index ω and an exponential cutoff for large values x. This transition occurs around the typical cell occupancy in overdense cells where $N = N_c$, i.e. $x = 1$.

Extending such studies of one-point quantities for the first time to their two-point analogues such as 2CPDF and 2VPF, we were able to check how overdense cells are biased with respect to the underlying mass distribution. We find, in accordance with hierarchical predictions, that the bias is an intrinsic property of collapsed objects. One can associate scaling variable x with every collapsed object, which is a function of cell occupancy N, radius of the cell, and the variance of matter distribution $\bar{\xi}_2$ on that length scale. We also found that the bias is factorizable and for two different objects with different values of scaling parameters x_1 and x_2 and with intrinsic bias $b(x_1)$ and $b(x_2)$, it is possible to express their bias with respect to each other as $b(x_1)b(x_2)$.

Elsewhere in the literature, extensions of the Press-Schechter theory have been used to compute the bias of collapsed objects (Mo & White 1996) and also their higher-order one-point moments such as S_N parameters (Mo et al. 1997). These are in agreement with the results we have obtained, especially those presented in Paper I. However, it remains unclear whether the generalisation of this formalism to two-point quantities (such as the CCs) will produce accurate predictions. Other works which have focussed on predicting bias and higher-order correlation functions are related to extensions of the Zel'dovich approximation (Lee & Shandarin 1997,1998, Catelan et al. 1997). It would be a very important test of the usefulness of these, and other approximation schemes for non-linear gravitational evolution to see if they can actually reproduce the multi-point statistics of collapsed objects as well as the simple hierarchical ansatz we have discussed here.

All these theoretical predictions were made in the lowest order of $(\xi_{ij}/\bar{\xi}_i)$, but were found to describe simulation results even when two cells are not separated by large distances.

Just as the VPF is related to the generating function of the S_N parameters, the 2VPF is related to the generating function of the cumulant correlators (CCs). In earlier numerical studies (Munshi & Melott 1998) we have shown that, in general, CCs are factorizable, so that $C_{pq} = C_{p1}C_{q1}$, when two cells are separated by a large distance. Such a decomposition of CCs is related to the factorization of the 2VPF which we have established in this paper, thus confirming our earlier findings. The

factorization property of bias is also a direct consequence of the scaling properties we study here, so our results on the subject of bias also vindicate our earlier findings.

Our numerical study was done in 2D and we plan to extend such analysis in case of 3D in near future, incorporating studies of higher-order correlations of over-dense cells.

It is interesting to note that although we have demonstrated that the hierarchical ansatz (62) is a very good approximation to the behaviour of higher-order correlation functions, stable clustering is known to be violated in 2D in highly non-linear regime (Munshi et al. 1998a). It is often argued that stable clustering is a necessary ingredient in the hierarchical ansatz. This appears not to be the case. We shall discuss this issue in more detail in a future paper.

Finite sample (or simulation) size known to introduce a sharp cut-off in the CPDF for $N = N_{\rm max}$ corresponding to the density of the densest densest cell in the catalog. Extending methods developed earlier (Munshi et al. 1998c) for correcting finite volume effects in the CPDF $P_l(N)$ by supplementing the missing information beyond $N = N_{\rm max}$ from scaling function $h(x)$ (which can then be used to compute the S_N parameters), we developed a similar method for MCPDF in this paper. At the level of 2CPDF e.g., one can use the scaling property of $b(N)$ which is encoded in scaling function $b(x)$ to correct or estimate finite volume effects in the extraction of CCs. We show that this technique can be generalised in a straightforward manner to the case of MCPDFs. At present there is no other general formalism to estimate the finite volume error for multi-point statistics. Our method provides the first step in this direction.

Acknowledgments

Dipak Munshi acknowledges support from PPARC under the QMW Astronomy Rolling Grant GR/K94133. Peter Coles received a PPARC Advanced Research Fellowship during the period when much of this work was completed. We are grateful for support under the NSF-EPSCoR program, as well as the Visiting Professorship and General Research Funds of the University of Kansas.

REFERENCES

Balian R., Schaeffer R. 1988, ApJ, 335, L43
Balian R., Schaeffer R. 1989a, A & A, 220, 1
Balian R., Schaeffer R. 1989b, A & A, 226, 373
Baugh C.M., Gaztañaga E., Efstathiou G., 1995, MNRAS, 247, 1049
Beacom, J.F., Dominik K.J., Melott A.L., Perkins S.P., Shandarin S.F., 1991, ApJ, 372, 351
Bernardeau F. 1992 ApJ, 392, 1.
Bernardeau F. 1994a, ApJ, 433, 1
Bernardeau F. 1994b, A&A, 291, 697
Bernardeau F. 1995, A&A, 301, 309
Bernardeau F. 1996, A&A, 312, 11
Bernardeau F., Schaeffer R. 1992. A&A, 255, 1.
Bond J.R., Cole S., Efstathiou G., Kaiser N., 1991, ApJ, 379, 440
Bouchet F.R., Hernquist L., 1992, ApJ, 400, 25
Bouchet F.R., Schaeffer R., Davis M., 1991, ApJ, 383, 19
Catelan P., Lucchin F., Matarrese S., Porciani C., 1997, (astro-ph/9708067)
Colombi S., Bouchet F.R., Schaeffer R., 1992, A& A, 263, 1
Colombi S., Bouchet F.R., Schaeffer R., 1994, A&A, 281, 301
Colombi S., Bouchet F.R., Schaeffer R., 1995, ApJS, 96, 401
Colombi S., Bouchet F.R., Hernquist, L., 1996, ApJ, 465, 14
Davis M., Peebles P.J.E., 1977, ApJS, 34, 425
Fry J.N., 1984, ApJ, 279, 499
Gaztañaga E., 1995, ApJ, 454, 561
Gaztañaga E., Frieman J., 1994, 437, 13
Goroff M.H., Grinstein B., Rey S.-J., Wise M.B, 1986, ApJ, 311, 6
Gramann M., 1992, ApJ, 401, 19
Hamilton A.J.S. 1988, ApJ, 332, 67
Jain B., Mo H.-J., White S.D.M., 1995, MNRAS, 276, L25
Jing Y.P, 1998, preprint (astro-ph/9805202)
Juszkiewicz R., Bouchet F., Colombi S. 1993, ApJ, 412, L9
Juszkiewicz R., Weinberg D.H., Amsterdamski P., Chodorowski M., Bouchet F., 1995, 274, 20
Kauffmann G., Melott A.L., 1992, ApJ, 393, 415
Klypin A.A., Shandarin S.F., 1983, MNRAS, 104, 891
Kofman L.A., Pogosyan D.Yu., Shandarin S.F., Melott, A.L. 1992, ApJ, 393, 437
Lee J., Shandarin S.F., 1997, ApJ, 500, 14
Lee J., Shandarin S.F., 1998, ApJ, 505, L75
Lokas E.L., Juszkiewicz R., Bouchet F.R., Hivon E., 1996, ApJ, 467, 1
Lucchin, F., Matarrese, S., Melott, A.L., & Moscardini, L., 1994, ApJ, 422, 430
Meiksin, A., Szapudi I., Szalay A., 1992, ApJ, 394, 87

Melott A.L., 1986, PRL, 56, 1992
Melott A.L., Shandarin S.F., 1993, ApJ, 410, 469
Melott A. L., Weinberg D.H., Gott J.R., 1983, ApJ, 328, 50
Mo H.-J., White S.D.M., 1996, 282, 347
Mo H.-J., Jing Y.P., White S.D.M., 1997, MNRAS, 284, 189
Maurogordato S., Lachieze-Rey M., 1987, ApJ, 320, 13
Maurogordato S., Schaeffer R., da Costa L.N., 1992, ApJ, 390, 17
Munshi D., Chiang L.Y., Coles P., Melott A.L., MNRAS, 1998a, 293, L68
Munshi D., Coles P., Melott A.L., 1998b, MNRAS, in press, (Paper I)
Munshi D., Melott A.L., 1998, ApJ submitted (astro-ph/9801011)
Munshi D., Bernardeau F., Melott A.L., Schaeffer R., 1998c, MNRAS in press (astro-ph/9707009)
Peebles P.J.E., 1980. The large scale structure of the universe (Princeton University Press, Princeton)
Porciani C., Matarrese S., Francesco, L., Catelan, P., 1998, MNRAS in press (astro-ph/9801290)
Press W.H., Schechter P.L., 1974, ApJ, 187, 425
Ruamsuwan L., Fry J.N., 1992, 396, 416
Schaeffer R., 1984, A&A, 134, L15
Scoccimarro R., Colombi S., Fry J.N., Frieman J.A., Hivon E., Melott A.L., 1998, ApJ, 496, 586
Szapudi I., Szalay A., 1993, ApJ, 408, 43
White S.D.M. 1979, MNRAS, 186, 145
Szapudi I., Colombi S., 1996, ApJ, 470, 131
Szapudi I., Dalton G., Efstathiou G.P., Szalay A., 1995, ApJ, 444, 520
Szapudi I., Szalay A.S., 1993, ApJ, 408, 43
Szapudi I., Szalay A.S., 1997, ApJ, 481, L1
Szapudi I., Szalay A.S., Boschan P., 1992, ApJ, 390, 350
Valageas P., Schaeffer R., 1997, preprint (astro-ph/9710128)
Yano T., Gouda N. 1998, ApJ, 495, 533